DOCUMENT RESUME

ED 258 809 SE 045 805

AUTHOR Mason, E. J.; And Others
TITLE Two Structured Interviews for Studying Development of Logical Reasoning in Children and Adolescents. Internal Report 79 ON 12.
INSTITUTION Catholic Univ., Nijmegen (Netherlands).
PUB DATE Aug 79
NOTE 52p.; A report of the Psychology Laboratory.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Cognitive Development; *Comprehension; Elementary Secondary Education; *Interviews; *Logical Thinking; *Measures (Individuals); Research Methodology; *Test Construction
IDENTIFIERS Netherlands

ABSTRACT

This paper describes the development and characteristics of structural interview procedures designed to aid researchers in investigating children's understanding of their logical arguments. The procedures comprise a system which assesses verbal expression, understanding of implication, and ability to reason logically; it also assesses understanding of parts of conditional arguments (the arguments presented as a whole) and of class inclusion aspects of implication. The first section of the paper describes the construction of the interviews and suggests modifications for their use in other kinds of studies. The second section describes the scoring methods used with the interviews. Inter-rater and scale realiability are also considered. The third and final section offers conclusions and suggestions for using the instrument. Illustrations of the interviews, scoring forms, and other related materials are provided in appendices. The instrument and procedures have been used successfully with subjects ranging from 5- to 15-years-old in the region of Nijmegen, the Netherlands. (JN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED258809

internal report 79 ON 12

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Two structured interviews for studying development of logical reasoning in children and adolescents.

E. J. Mason
V. J. E. Haars
P. J. M. Dingemans
P. Bierman - van Duuren

BEST COPY AVAILABLE

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Emanuel J. Mason
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

SE 045 805

Vakgroep Ontwikkelingspsychologie
Psychologisch Laboratorium
Katholieke Universiteit Nijmegen
the Netherlands 2

TWO STRUCTURED INTERVIEWS FOR STUDYING DEVELOPMENT OF LOGICAL REASONING IN CHILDREN AND ADOLESCENTS[1)]

E.J. Mason[2)]
V.J.E. Haars
P.J.M. Dingemans
P. Bierman-van Duuren

August, 1979.

1) This research was partially supported by a visitor's grant to the first author from the Netherlands Organization for Pure Scientific Research (Z.W.O.).

2) On leave from the University of Kentucky.

KATHOLIEKE UNIVERSITEIT - ERASMUSLAAN 16 - NIJMEGEN - THE NETHERLANDS

APPENDICES

1. Pictures placed on cards of: a) Foods
   b) Football
   c) Tree
   d) Trays containing blocks and animals.

2. Practice questions and general instructions to subjects.

3. Sample question from Form PW.

4. Samples questions from Form CA.

5. Sample of instructions for examiner (Form PW).

6. Instructions for examiner (Form CA).

7. K-R coding Form.

8. Scoring instructions for Form C-L.

9. Form C-L for CA.

10. Form C-L for PW.

INTRODUCTION

Logical reasoning is considered by many to be fundamental to mature thinking (see, for example, Ennis, 1975). As such, logic has been suggested as a basis for educational programs, particularly in mathematics (e.g., Nuffield Foundation, 1972). Further, formal instruction in logic and critical thinking has been identified as an important component in the education of a child (Ennis, 1975; Ennis & Paulus, 1965; Hadar, 1975). The way that logical thinking develops has been a major interest of developmental psychologists for the past few decades. stimulated largely by the work of Jean Piaget (1925, 1963, Inhelder & Piaget, 1958, 1964).

The mechanics or processes of thinking or behaving according to logical rules, and their development are still not clearly understood. Explanations provided by Piaget and his followers do not sufficiently clarify the issue. For example, many writers have criticized the logic of Piaget as illogical (Ennis, 1975; Osherson, 1975; Parsons, 1960), unclear (Brainerd, 1978; Flavell, 1977), or not supported by empirical data (Bynum, Thomas & Weitz, 1972; Weitz, Bynum & Thomas, 1973). The content area in which the person tries to reason can also afflict the ability to perform logically (Nolen, 1976; Piaget, 1972).

Of the various approaches that have been used to study the development of logic, each has posed unique difficulties. For example, some have used formal syllogisms to explore children's reasoning. However, this kind of logic may not correspond to the processes of thinking (Braine, 1978; Henle, 1962). A rather thorough discussion of this is provided by Osherson (1975, p. 16-25). Further, much of this research has involved the use of group paper-and-pencil examination and required respondents to indicate whether the argument forms are valid or not (e.g. Hadar, 1975; Roberge, 1970; Roberge & Mason, 1978; Shapiro & O'Brien, 1970).

The group administered paper-and-pencil approach is useful for testing large numbers of subjects, but does not offer an opportunity for the researcher to explore the meaning of the logic with the child. It would be difficult, for example, using this approach to determine whether

subjects are giving correct answers to logical arguments because they have memorized an algorithm, have a partial understanding that is just enough to give the correct response, or have a complete understanding of the logic underlying the argument. This can be shown simply using the following syllogism:

If all men are mortals
and Socrates is a man.
Then Socrates is a mortal.

A child who is asked about the validity of this argument might respond affirmatively because he or she understands the premises in any of the following ways:

a) There are men and there are mortals.
b) There are mortals and there are men.
c) Men must be mortals but mortals do not have to be men. Since Socrates is a man, he must also be a mortal.
d) To be mortal implies being a man also. For Socrates to be a man he must be a mortal.

Although all of the understanding listed above would probably lead a child to say that the argument is correct, (c) shows the most complete understanding, and (d) is clearly incorrect. The remaining understandings are not complete, although they would lead the child to respond correctly.

Individual interviews have been used to explore children's understanding of logic with some success (e.g., Falmagne, 1975; Inhelder & Piaget, 1958; Kodroff & Roberge, 1975; Kuhn, 1977). There are difficulties with interviews, however, that are less troublesome with paper-and-pencil group administered instruments. These problems involve administration and scoring. Administration problems can be reduced by training interviewers not to influence the behavior of the respondent other than to encourage responses. Difficulties in scoring and interpreting interview data can also be reduced by training scorers and providing well-defined guidelines for interpreting and classifying responses.

The present paper describes the development and characteristics of a structured interview procedure designed to investigate children's understanding of their logical arguments. The first section describes the construction of the interviews and suggests some modifications in order to use them for other kinds of studies. The next section describes the scoring methods devised for use with the interviews. Inter-rater and scale reliability are covered next. Finally, conclusions are drawn, and suggestions are made for using the instrument. Appendices are provided which contain illustrations of the interviews, scoring forms, and other materials.

DESIGN OF THE INTERVIEWS

The interviews were designed to explore two different kinds of questions. The first was specifically designed to explore the child's understanding of the premises and conclusions of syllogisms, the second to explore the child's understanding of the syllogism in total. Although the interviews are described here with specific content and materials for each question, these could easily be altered to accomodate different research needs and questions.

The two interviews were based on four syllogism forms: Modus Ponens (MP) ($P \supset Q, P \therefore Q$), Modus Tollens (MT) ($P \supset Q, \bar{Q} \therefore \bar{P}$), Affirmation of the Consequent (AC) ($P \supset Q, Q \therefore P$), and Denial of the Antecedent (DA) ($P \supset Q, \bar{P} \therefore \bar{Q}$)[3]. Truth table analysis will reveal that in the forms used, MP and MT are valid arguments while AC and DA are invalid.

The two forms are described in the following sections. Since the first form was designed to compare understanding of premises individually and in the context of a whole argument, it is identified as the Part-Whole Form (Form PW). The second was only designed to study the meaning of the whole argument, and is therefore called the Complete Argument Form (Form CA).

---

3 A line over a letter means negation (i.e., $\bar{P}$ should he read "not P"), the symbol $\supset$ means implication ($P \supset Q$ should be read "P implies Q"), and the $\therefore$ symbol means the conclusion follows, and should be read "therefore".

FORM PW

This form was developed to explore relationships between understanding of the superordinate and subordinate clauses in the major premise and logical processing of the whole argument. As part of the understanding of the conditional implication, the child's understanding of the class relationships in the content of the premises was explored. Further, the form was designed to explore two different content dimensions at the same time, in the case of the illustration discussed here, concrete and verbal. Since there was a possibility that experience with a partial presentation might influence a child's performance on the total argument, and vice versa, the interview was balanced for presentation of part first or whole first, and for the two content dimensions producing four parts as follows:

Part I - Whole first, concrete.
Part II - Part first, concrete.
Part III - Whole first, verbal.
Part IV - Part first, verbal.

Items were similar to the three-choice response paradigm used in other reasoning research (Ennis & Paulus, 1965; Roberge, 1970). That is, subjects were required to state whether they thought the logic was valid or invalid with the following three choices: Yes (it is valid), No (it is not valid), Maybe (it may be or may not be valid depending upon circumstances). However, in the present interviews subjects were asked to give reasons for their answers.

In each concrete presentation the examiner used two trays (clear plastic 18.0 cm x 11.6 cm x 3.2 cm). One had ten small plastic animals in it (varying sizes ranging in height from about 2 cm to about 5.5 cm and including one each of a lion, deer, antelope, rhinocerous, elephant, kangeroo, black cheeta, white polar bear, moose, and a white dog; the dog statue came from the standard 1959 Stanford-Binet Intelligence Test Kit.). The other tray (identical to the first) contained 12 green blocks ($2.5\ cm^3$) from the Stanford-Binet Kit. In the animal tray the dog was placed separately from the cluster of other animals at one end of the tray. Further, separate from the trays and outside of them, the examiners were

given an unpainted wood block of 2.5 cm x 2.5 cm x 8.0 cm, and a second dog identical to the one in the animal tray.

For the verbal presentation, cards (16 cm x 10 cm) containing drawings of a tree and a football (soccer ball) were used. Illustrations of the materials described here are in the appendix.

The format of the whole-first presentation (Parts I and II) for each item is illustrated below using an example from the animals-dog concrete content[4].

1. Examiner puts tray A before the subject and says[5]:

IF ALL DOGS ARE ANIMALS	(as words are spoken, E points to dogs and then rest of animals).
AND THIS IS A DOG	(E shows S the larger statue of a dog that had been kept out of sight).
THEN, I SHOULD PUT THIS WITH THE ANIMALS	(E places dog in box near dogs).

A. DOES THAT CONCLUSION FOLLOW?

If the child does not give a standard answer, (yes, no, maybe) E says:

REMEMBER, YOU CAN ONLY SAY YES, WHICH MEANS IT MUST FOLLOW:
NO, WHICH MEANS IT CANNOT FOLLOW, OR
MAYBE, WHICH MEANS IT MAY BE TRUE SOMETIMES, BUT YOU HAVE NOT BEEN TOLD ENOUGH TO BE CERTAIN.

When S gives a standard answer, E asks:

B. WHY DO YOU SAY THAT?

After S's answer, E continues:

C. TELL ME WHAT IT MEANS TO YOU WHEN I SAY,
ALL DOGS ARE ANIMALS.

When S responds, E asks:

D. ARE THERE MORE DOGS OR ANIMALS?

When S responds, E continues, NOW TELL ME WHAT THIS MEANS:

---

[4] The reader can easily construct the remainder of the items by substituting the "cubes and blocks", "trees and plants" and "balls and round things" in the premises of the four syllogism forms. Table 1 should be of some help here.

[5] The words the examiner says are in upper case letters, instructions

E. THIS IS A DOG (E points to larger dog in tray)
THEN, I SHOULD PUT THIS WITH THE ANIMALS.

The following is an illustration of an item from the part-first presentation (Parts II and IV). Again the animals-dog content is used. There were also items with blocks-cubes, trees-plants, and balls-round things content.

4. Examiner puts tray A before the subject and says:

A. TELL ME WHAT IT MEANS TO YOU WHEN I SAY,
ALL DOGS ARE ANIMALS.

When S responds, E asks:

B. ARE THERE MORE DOGS OR ANIMALS?

When S responds, E continues, NOW TELL ME WHAT THIS MEANS:

C. THIS IS NOT AN ANIMAL (E holds up larger dog)
THEN, I SHOULD PUT THIS WITH THE DOGS.

After S responds, E says,
NOW LISTEN TO THIS ONE.
IF ALL DOGS ARE ANIMALS (E points to dogs than animals as words are spoken)
THIS IS NOT AN ANIMAL
THEN, I PUT THIS WITH THE DOGS.

D. DOES THAT CONCLUSION FOLLOW?

If the child does not give a standard answer, (yes, no, maybe), E says:
REMEMBER, YOU CAN ONLY SAY YES, WHICH MEANS IT MUST BE TRUE.
NO, WHICH MEANS IT CANNOT FOLLOW. OR -
MAYBE, WHICH MEANS IT MAY BE TRUE SOMETIMES. BUT YOU HAVE NOT BEEN TOLD ENOUGH TO BE CERTAIN.

When S gives a standard answer, E asks:

E. WHY DO YOU SAY THAT?

The content by syllogism matrix in Table 1 shows the schematic used to construct the items. There were two items in every part for each kind of content and syllogism. Since there were four syllogisms, and two kinds of content crossed in each part, there was a total of eight items in a part, and 32 items in the whole interview.

Table 1

Content matrix for constructing test items for Form PW [a)]

syllogism form	*Concrete Content*				*Verbal Content*			
	Part - First		*Whole - First*		*Part - First*		*Whole - First*	
	Dogs/ Animals	Cubes/ Blocks	Dogs/ Animals	Cubes/ Blocks	Trees/ Plants	Balls/ Round things	Trees/ Plants	Balls/ Round things
MP	1	5	1	5	1	5	1	5
MT	2	6	2	6	2	6	2	6
DA	3	7	3	7	3	7	3	7
AC	4	8	4	8	4	8	4	8

[a)] Numerals in cells of table used to randomly order items.

After the items for each part were constructed, they were randomly arranged for presentation. This was accomplished by numbering the items consecutively and then choosing them in random order for placement in the interview booklet using a table of random numbers. Each part required 10 to 15 minutes for administration. The interview schedule for Form PW is reproduced partially in the Dutch language in the appendix.

FORM CA

The second interview schedule was designed only to explore children's understanding of complete arguments. Again, the MP, MT, DA, and AC argument forms were the basis for the instrument, however, other forms could be added. In the present study, there were two forms of each argument (Roberge, 1971) for the purpose of investigating certain specific research questions. These questions will be the topics of later reports, and will not be covered here.

In testing Form CA, each conditional argument was illustrated using the 2.5 cm^3 green blocks from the Stanford-Binet Intelligence Scale Kit,

and the following other wooden blocks:

2-yellow cubes (2.5 cm^3)

2-green circles (diameter 3 cm, height 1 cm)

2-yellow circles (diameter 3 cm, height 1 cm)

1-brown rectangular carpet remnant (30 cm x 37 cm).

The carpet remnant was divided into two sections by a 1.5 cm wide strip of bright red tape placed 10 cm from the top edge, and from one 37 cm side to the other. These blocks, circles, and carpet were used to test the interview format, and may be altered or replaced with other materials as other research questions warrant.

Each item in Form CA used the following format.

Examiner (E) places carpet before the child, puts on it : (some blocks) and says,

IF ALL BLOCKS ARE WOOD

AND THIS IS ..... (E holds one block apart from the others)

THEN IT IS .....

A. E then says: DOES THAT FOLLOW?

If the child does not give a standard answer, E says:

REMEMBER YOU CAN ONLY SAY YES, WHICH MEANS IT MUST BE TRUE,

NO, WHICH MEANS IT CANNOT BE TRUE, OR

MAYBE, WHICH MEANS YOU HAVE NOT BEEN TOLD ENOUGH TO BE CERTAIN.

B. When the child has given an answer, E asks:

WHY DO YOU SAY THAT?

From the argument (four types in two forms each), shape of the minor premise block (circular or cube), and color (yellow or green), a total of 32 items were constructed. This is illustrated in Table 2. As with the last form, the items were ordered randomly in the interview. In addition, the 32 items were divided in half, so that 16 of the items could be administered typically in an administration. This would require the child to attend for periods of 10 to 20 minutes, usually about 12 minutes.

Table 2

Content Matrix for forming Test Items for Form CA[a)]

Syllogism form	*circles*		*cubes*	
	yellow	green	yellow	green
MP 1.	1	9	17	25
2.	2	10	18	26
MT 1.	3	11	19	27
2.	4	12	20	28
DA 1.	5	13	21	29
2.	6	14	22	30
AC 1.	7	15	23	31
2.	8	16	24	32

a) Numerals in cells used to randomly order item

For research questions not in the scope of the present discussion, half of the subjects were administered the items with the content of the minor premise placed separately from the remainder of the blockson the carpet using the red tape strip as a barrier. The rest of the subjects were shown the carpet with content of the minor premise included with the other blocks. This aspect will not be discussed further here.

ADMINISTRATION

The interviews were administered during the regular school day individually by graduate students in developmental psychology from the University of Nijmegen. Interviews were recorded on cassette tape for later analysis. Examiners had specific instructions describing what to say. Copies of these instructions are included with the interviews in the appendices of this report.

Children and adolescents were interviewed individually in a quiet room outside of their regular classroom. The setting varied slightly in each school, however, generally the examiner sat at a table directly across from the child with the test materials arranged between them and the cassette tape recorder (with built-in microphone) off to one side.

Each session was designed to take no more than 20 minutes. For Form PW, generally one or two parts were administered during a single session. With Form CA, half the items were typically administered in a session, however, with older or more able subjects, occasionally all items could be administered within the 20 minutes session. The order of administration of the parts for Form PW, and the halves for Form CA was randomly determined.

Each child's interview was preceded by a series of instructions and some practice examples. The practice examples were devised to give the child some experience with the interview format and the use of the terms YES, NO and MAYBE in responding to the questions. The practice problems did not contain any of the formal arguments used in the interview. The practice examples and accompanying instructions are included in the Appendices.

SUBJECTS

Subjects for the studies described here were students in nursery, primary, and secondary schools in the region of Nijmegen, the Netherlands. For testing Form PW, kindergarten, 2nd, 4th, and 6th grade children were randomly sampled as were students from the second year of secondary school. Secondary school students were selected in the same ratio for HAVO, MAVO, and VWO as revealed by national statistics on secondary schools (Centraal Bureau voor de Statistiek, 1976/77). Thus,2nd year secondary school children were chosen from a community secondary school serving the region of Nijmegen in the ratio of 7:2:3 (MAVO to HAVO to VWO).

For testing Form CA, students were sampled randomly from two elementary schools and a secondary school in Nijmegen. The elementary students were in grades 1, 3, and 5, and the secondary students were in the first grade of secondary school. Since there is no distinction between HAVO, MAVO and VWO at that level, no striated sampling could be done, although it should

be stated that at the particular secondary school which supplied the sample, a larger portion of the students go on to a VWO program than is true nationally for children of that age.

Table 3 summarizes the characteristics of the students.

Table 3

Characteristics of Students who participate in Studies

Grade	*Number of boys*	*Number of girls*	*Average age (yrs-months)*
Kindergarten	5	5	6-0
1	7	5	7-0
2	6	5	8-2
3	6	6	9-1
4	6	6	10-2
5	6	6	11-2
6	6	6	12-3
Secondary 1	9	9	13-0
Secondary 2	6	5	15-8
	57	53	

TABULATING AND SCORING

In this section three stages in the processing of the data collected in the interviews are discussed. The first is the transcription of the data from the cassette tapes to handwritten form. This was done to facilitate scoring, and to make a record of each interview. The second stage involved the actual scoring of the child's assessment of the validity of the syllogism and the related class inclusion question. The third stage included the classifying of the child's verbal explanations of his or her reasoning into categories of verbal and logical explanations.

TRANSCRIPTION

All interviews were recorded using a cassette tape recorder (Philips, N2215 automatic portable cassette). The recordings were made so that there would be an accurate record of each interview. However, it was felt that scoring from the recorded interview would be difficult because responses to specific questions cannot be addressed easily on a cassette tape. Also, the time required to play a tape in order to listen to one testing session would be greater than the time needed for reading a transcript of the session if it were in written form. For that reason, the children's responses were transcribed. Forms were used which were designed to facilitate scoring and at the same time reduce the amount of writing necessary to record the essence of the child's responses.

For Parts I and III of Form PW in which the whole syllogism was presented first, responses were transcribed using the following format for each item:

A.	YES	NO	MAYBE
B.			
C.			
D.	P		Q
E.			

The transcription format used with each item of Parts II and IV of Form PW was:

A.			
B.		P	Q
C.			
D.	YES	NO	MAYBE
E.			

Note: in both formats that the letters correspond to the order of the questions asked in each Part.

Transcribers used these forms in the following way. They circled the child's response to the question about whether the argument was valid (YES, NO, or MAYBE), and to the class inclusion question (subordinate (P) or superordinate (Q)). For the remaining sections, the transcribers were told to record in the child's words what they thought were the essentials of the child's response. They were told to omit any irrelevant verbalization, and to keep in mind that they would be required to score the child's understanding of his or her response on the basis of what was written in the transcripts. Since there was thus an element of judgment required on the part of the persons doing the transcription, the question of inter-transcriber reliability was raised. This question was investigated and will be discussed further in the section on reliability.

The format for transcribing responses to each item in the Form CA interviews is shown below.

A.	YES	NO	MAYBE
B.			
C.			

In the section labelled "A" the transcriber circled YES, NO, or MAYBE to indicate the child's response. In section B, the essentials of the child's response were noted. Finally, in section C, the time elapsed in seconds from when the examiner began the question to the moment when the child responded with "YES", "NO", "MAYBE". The time was recorded using a hand-held stopwatch with a sweep second hand, and was recorded to 1/10 second accuracy. The reliability of the timing recorded by transcribers was studied as well as the reliability of the transcription and will be discussed later.

SCORING

In Form PW two questions in each item, and in Form CA one question in each item could be scored directly. That is, the items for which the child was responding with a one word answer that was either correct or incorrect could be scored directly without further interpretation. These were scored using a key. An incorrect answer was credited as "0" and a correct one as "1". The total score was the sum of the 1's and 0's.

CLASSIFICATION OF VERBAL EXPLANATIONS

Two instruments were used for classifying verbal responses. One, the Kodroff-Roberge system, focused on categories of verbal explanation based on the technique used with some success to classify children's verbal reasoning (Antonak & Roberge, 1978; Kodroff & Roberge, 1975). The second classification system (C-L) was directed more at the class and logical relationships expressed by the children. Copies of the forms designed for these two classification systems may be found in the Appendices. The specific instructions given to the scorers are also included in the Appendices. Scorers practiced the two systems until they felt that they were familiar enough with them to use them. Then inter-rater reliability was determined. These reliabilities are discussed in the reliability section.

KODROFF-ROBERGE SYSTEM

The K-R system used in the present studies was adapted somewhat from the instrumentation of Kodroff and Roberge. To begin with, the original system had six categories of verbal behavior. The first four categories were used as proposed by Kodroff and Roberge. They included <u>verbalization deficits</u> (an inability to discuss the logic behind a deficit, e.g., "I don't know"), <u>memory deficits</u> (repeats a premise incorrectly, loses

train of thought or becomes confused, e.g., "Because this is a plant and this ... I forgot"), arbitrary explanations (brings in facts not given in the premises, e.g. "I had a dog like that once and he was an animal"), and structural deficits (response indicates an awareness of conditional logic but not adequate conditional reasoning, e.g. "There are dogs and there are animals"), patterned explanations, the sixth K-R category was also used as originally proposed (response indicates that the premises are organized in a pattern, at least three terms in the item are included ("e.g., Dogs and animals. This is a dog and an animal"). The fifth category, conditional reasoning was changed slightly. In the current system a response was only scored as conditional reasoning if the response showed some understanding of conditional reasoning that was incorrect. The child must have included the words "If" and "then" in the response (e.g. If this is a plant, then it must be a tree").

Two other changes were made in the original version of the K-R scoring system for the present studies. One was the addition of a category called other for verbal explanations which did not fit the other categories. The second was the addition of a category entitled good logic. To be classified as good logic the response must not have fit in any of the other categories, must have shown mature interpretation of the premises and the implications involved, and the proper relationships between the superordinate and subordinate terms in the major premise (e.g."Since all dogs are animals, there can be no dogs that are not animals. Therefore, this is a dog, so it is also an animal").

C-L CODING FORM

The C-L coding form was devised to identify the kinds of logic represented in subjects' responses. There were two kinds of ratings made of each response using this system. The first kind was classification relationships. The second was logical relationships.

Classification relationships referred to the subordinate and superordinate characteristics of the terms in the premise. For example, in the premise "If all trees are plants", trees are part of the group of

things called plants. However, the child's verbalization may reflect on understanding of the premise with trees and plants as separate and equal concepts, or some trees as plants but others not, etc. These kinds of relationships were coded on the class relationships side of Form C-L. To illustrate, "If all trees are plants", may be understood by the child to mean any of the following:

a) There are equal numbers of trees and plants (P=Q).
b) There are more trees than plants (P>Q).
c) There are less trees than plants (P<Q).
d) Trees and plants both exist as separate unrelated concepts((P) (Q)).
e) Trees and plants are parts of the same concept ((PQ)).
f) Some trees are plants and some plants are trees ((P Q)).
g) Trees are part of the larger group of plants ((P) Q).
h) Plants are part of the larger group of trees ((Q) P).

These categories of meaning formed the class relationship categories of the C-L coding form.

The logical ratings on the other hand were based on Piaget's suggestions about the components of logical implication (see Ennis, 1975; Inhelder & Piaget, 1958). To these were added a few categories that the authors felt might be important. The following categories were checked when evidence for them could be found in the child's response:

a) P : The subordinate (P) clause or term is affirmed.
b) Q : The superordinate (Q) clause or term is affirmed.
c) $\bar{P}$: The negation of P is affirmed.
d) $\bar{Q}$: The negation of Q is affirmed.
e) P.Q : Both P and Q are recognized.
f) $\bar{P}$.Q : The negation of P and Q are affirmed.
g) $\bar{P}.\bar{Q}$: The negation of both P and Q are affirmed.
h) P.$\bar{Q}$: The existence of P without Q is affirmed.
i) N(P.$\bar{Q}$): The existence of P without Q is denied.
j) $\begin{matrix} P \rightarrow R \\ R \rightarrow Q \end{matrix}$: Evidence of transitive reasoning (from P to R and R to Q) is given.

Logical relationships were checked if they were implied by the child's response. For example, if the response was "When these are dogs, these are

animals because dogs are animals", the item would be scored as P.Q (where P represents the subordinate and Q the superordinate categories in the premises). But also, since it is implied that not all animals are dogs, or that dogs constitute a subordinate group of the category of things known as animals, the logical relationships $P.\bar{Q}$ and $\bar{P}.\bar{Q}$ could be inferred.

These kinds of inferences on the part of the scorers required that the scorers be trained. Further, in such a procedure, inter-rater reliability should be established to demonstrate that different raters can rate a protocol with some consistency, or to put it another way, that different people can see the same things in a protocol.

The logical and the class sides of the C-L coding form contained a column for "other", which was used when a response could not be placed in one of the categories. Scorers placed a check mark in the appropriate column for each of the class and logical relationships that could be inferred from a response.

Scorers were trained to use coding Form C-L by coding two or three protocols together after they had read and discussed. When they felt they were agreeing substantially, they rated the reliability tapes, and then the remainder of the protocols.

RELIABILITY

Reliability refers to the degree to which a test or procedure yields consistent information. For a test to be reliable, it must give the same score over repeated administrations or when equivalent versions are used, the scores for each individual should be close if the test is reliable. Thus, the higher the reliability the more dependable the results of the procedure.

In regard to the procedures discussed in the present report, reliability was a primary concern. The present section discusses reliability estimations of various types for each of the procedures used. The reliability of the transcription procedures are discussed first. Then the reliabilities of the total reasoning scores obtainable from Forms PW and CA are described. Following this, the Kodroff-Roberge verbal classification

system is discussed from the point of view of reliability of the raters. Finally, reliability estimations for the C-L coding Form are presented. These reliability studies were made using the same three third year graduate students in developmental psychology at the Catholic University of Nijmegen, who collected the data, then transcribed and recorded it.

A. RELIABILITY OF TRANSCRIPTIONS

Form PW - A random sample of five cassette tapes was chosen from a pool of 112 taped interviews, one tape from each of the five grades tests. Each cassette contained two parts of the four part form. Two of the tapes chosen contained parts I and II, the remainder contained Parts II and IV. The raters wrote an average of about 12 to 15 words per response for the 120 responses they each transcribed on the five tapes, as can be seen in Table 4. There were not significant differences between the raters in the number of words they transcribed. Further, the average correlations between the three raters for the number of words they transcribed, broken down by question and Part in Table 5, suggest a high degree of consistency between raters in terms of the number of words transcribed.

However, agreement in the number of words does not suggest agreement in the meaning of the words or the words used. The degree to which the raters agreed upon the meaning of their transcriptions was investigated by comparing their ratings using the K-R and C-L scoring systems, and is discussed later in this report. In regard to the degree to which there was agreement in the actual words used, and the order in which they appeared in the transcripts, a count was made with each rater paired with every other one. As can be seen in Table 4, an average of about 10 to 13 words appeared in the same order for each pair of raters.
Further analysis revealed correlations between pairs of raters ranging from .86 to .99 suggesting a high degree of consistency (see Table 6).

It was concluded on the basis of the data that there was a high degree of consistency in transcribing from the recorded interviews using the instructions for the raters that they should write only the portion

Table 4

Means and standard deviations of the number of words tabulated and the number of words agreed upon by three transcribers on 120 separate transcriptions for each part on Form PW.

	RATER					
	1		*2*		*3*	
Part	Mean	s.d.	Mean	s.d.	Mean	s.d.
1	13.33	7.04	14.67	8.145	13.80	7.77
2	13.18	9.80	12.98	7.66	12.23	6.75
3	12.94	7.66	14.31	9.53	13.49	7.60
4	11.58	8.64	12.08	8.67	11.52	9.92
	PAIRS OF RATERS					
	1 and 2		*1 and 3*		*2 and 3*	
Part	Mean	s.d.	Mean	s.d.	Mean	s.d.
1	12.25	6.03	12.20	6.07	12.62	6.59
2	11.17	5.84	11.16	5.84	11.68	7.09
3	12.37	7.23	11.88	6.09	12.63	6.85
4	10.63	7.54	10.08	6.80	10.13	6.64

of the child's verbal response that was pertinent to understanding how the child understood the logical argument, and his or her reasons for each answer.

Form CA - A random sample of eight taped interviews was chosen from a pool of 54 interviews. Two interviews were chosen from each of the first grade, third grade, fifth grade, and first year secondary school (seventh year of formal education).

Table 5

Correlations of the number of words transcribed in each Part in five sample tapes for the three questions in Form PW.

		QUESTIONS[a]					
		B (I, III) E (II, IV)		*E (I, III) C (II, IV)*		*C (I, III) A (II, IV)*	
Part	Transcriber	2	3	2	3	2	3
I	1	.95	.77	.98	.97	.94	.94
	2		.94		.97		.95
II	1	.92	.86	.82	.72	.96	.97
	2		.97		.75		.98
III	1	.94	.85	.96	.95	.97	.79
	2		.86		.99		.98
IV	1	.92	.79	.98	.98	.95	.92
	2		.89		.99		.98

[a] Part numbers are indicated in parentheses.

Table 6

Correlations of the number of words agreed upon between transcribers in five sample tapes for the three questions in Form PW.

		QUESTIONS[a]					
		B (I, III) E (II, IV)		*E (I, III) C (II, IV)*		*C (I, III) A (II, IV)*	
Part	Transcriber	1-3	2-3	1-3	2-3	1-3	2-3
I	1-2	.99	.93	.98	.98	.97	.94
	1-3		.94		.98		.93
II	1-2	.97	.93	.96	.91	.96	.94
	1-3		.96		.93		.96
III	1-2	.95	.97	.99	.96	.97	.94
	1-3		.97		.97		.98
IV	1-2	.93	.97	.99	.98	.97	.93
	1-3		.97		.98		.98

[a] Part numbers are indicated in parentheses.

Table 7 shows the mean number of words written by each transcriber, the mean number of words agreed upon in the same order of appearance for pairs of raters, and the mean times recorded for responses using a hand held stopwatch. As can be seen in the table, the second transcriber had fewer words recorded than the others. The difference in number of words between

Table 7

Means and standard deviations of the number of words written by transcribers, number of words agreed upon in order of appearance for pairs of transcribers, and the time recorded by transcribers with a hand held stopwatch.

Transcriber	*Number of words* *Mean*	*s.d.*	*Time (in seconds)* *Mean*	*s.d.*
1	19.77	15.36	12.62	7.60
2	12.54	8.00	12.51	7.66
3	18.23	13.49	12.74	7.50
	Number of words agreed upon in order *Mean*	*s.d.*		
Pairs of transcriber				
1 with 2	11.97	7.26		
1 with 3	17.00	12.05		
2 with 3	11.93	6.91		

this rater and the other two was statistically significant ($t_{\alpha/2}$= 5.13 and 4.26, respectively, $\underline{df}$ = 510, $\underline{p}<.001$). However, this lack of agreement between one of the raters and the other two does not suggest that the shorter transcriptions were necessarily less adequate than the longer ones for extracting meaning. This would be revealed in the application of the C-L and K-R scoring systems to the data if it is the case.

Table 8 shows the degree to which the number of words transcribed correlated among the three transcribers. Correlations were from .70 to .96 suggesting that even thcugh there was a difference in the number of words recorded among observers, there was moderate to high agreement on the items for which more words were required. Similar correlations were recorded for the correlations among pairs of raters.

Table 8

Correlations of total numbers of words transcribed, and number of words agreed upon in transcriptions by three recorders for Form CA.

	Total number of word		
Rater	*1*	*2*	*3*
1	1.00	.70	.96
2		1.00	.77
3			1.00
	Words agreed upon		
Pairs of raters	*1 with 2*	*1 with 3*	*2 with 3*
1 with 2	1.00	.78	.99
1 with 3		1.00	.77
2 with 3			1.00

Table 9

Correlations for the times recorded by three raters using Form CA.

Rater	*1*	*2*	*3*
1	1.00	.96	.96
2		1.00	.96
3			1.00

In tables 7 and 9, it can be seen that there was substantial consistency in time ratings across observers. This finding suggests that response latencies can be measured accurately by observers from a tape recorded interview using a hand held stopwatch.

B. RELIABILITY OF REASONING SCORES

Each form PW and AC was capable of yielding a total reasoning score. This total score was computed by totalling the correct responses YES, NO, or MAYBE for the question about whether each argument was valid. When the child was correct, a score of 1 was assigned to the item, and when the response was incorrect, a score of zero was assigned to the item. The total reasoning score thus represented a summation of all the 1s and 0s. The internal consistency approach was used to investigate the reliability of this total score.

Form PW - The total score was divided into the score for odd numbered and even numbered parts. Thus, each part contained two presentations of each of the argument formate. Two internal consistency formulas were used, the Spearman-Brown, and the one Cronbach (1960) suggests as a more adequate method:

$$r_{tt} = 2\left(1-\frac{s_a^2+s_t^2}{s_t^2}\right)$$

Where: r_{tt} = the reliability coeficient

s_a^2 = the variance of one half of the rest

s_b^2 = the variance of the second half, and

s_t^2 = the variance of the total test.

Both formulas yielded nearly identical reliability estimates of .88 (.878 for the Cronbach, and .880 for Spearman-Brown). This is a moderately high reliability coeficient, and would suggest that the total score obtainable from Form PW is internally consistent.

Form CA - The total score for Form CA was divided into two parts by assigning two administrations of each argument form to one of two parts. The same internal consistency approaches were used to estimate reliability as were used with Form PW. Again almost identical estimates were obtained of about .93. (.932 for Cronbach, and .933 for Spearman-Brown). This reliability coefficient was considered rather high and suggested internal consistency for the total reasoning score obtainable from Form CA.

C. RELIABILITY OF THE REVICED KODROFF-ROBERGE SYSTEM

It will be remembered that the Kodroff-Roberge system was a method for classification of verbal responses to questions requiring reasoning. Therefore, the reliability question with this system concerns the consistency with which different raters use the categories of response. Raters made their ratings from the transcriptions of the tapes. Each rater used his or her own transcription since it was felt that in a practical scoring situation, the rater would use his or her own transcription. Furthermore, the transcribers were given the instructions to only transcribe the parts of the child's verbal response that they felt were necessary to show the child's understanding of the reasoning principles it used. Therefore, the rater's classifications on this and the C-L coding forms could be considered evidence of content validity of the transcripts if the inter-rater reliability is high.

The determination of inter-rater agreement with categorical data is somewhat more complex than simply counting the number of times two or more raters agree and the number of times that they do not. Many of the problems involved have been discussed by Cohen (1960), Light (1971), and Robinson (1957). Basically, the problems involve a distinction between demonstrating patterns of agreement versus patterns of association. Without going into greater detail here, the methods suggested by Light (1971) and Cohen (1960) were used with the present data.

For each of the Forms, PW and CA, a three dimensional matrix of agreement was constructed. There was one dimension for each rater. The eight categories of ratings were placed along each dimension to form an

8x8x8 cube containing 512 "cells". Using these cubes, coefficients of agreement were calculated for all three raters using Light's procedure (1971, p. 365-370), and for pairs of raters using Cohen's (1960) method.

Form PW - The overall coefficient of reliability (Km) of the ratings given by the three raters was computed to be .74 with the Light formula. Inter-rater reliabilities (K) were computed to be .75, .77, and .70 respectively for pairs of raters. These reliabilities were considered moderate and were statistically significant ($p<.05$). It should be noted that there was somewhat greater agreement between raters for the responses of children who were younger (Kleuterschool and second grade) and those who were older (second year of secondary school). At those ages raters said the children seemed better able to express their understanding, or lack of it. The fourth and sixth graders may have been in a transitional stage between more mature logic and less mature, and/or between more and less mature expressive ability. This, of course, is a broader research question than should be discussed here, and will be the subject of later reports. However, it should be recognized that if at some ages children are not sure what they mean to say about their logic, or are not sure how to explain it, this could effect the consistency of those trying to rate what they mean.

In addition the percent of agreements between pairs of raters were computed. These ranged between 78.13% and 83.75%. This was considered substantial, and to reflect further on the consistency between raters. A total of 160 responses was rated in the five tapes by the three raters.

Form CA - The overall coefficient of reliability (Km) for the three raters was computed to be . 0. For pairs of raters inter-rater reliability estimates ranged from .66 to .74 using Cohen's method. Further, the pairs of raters agreed on from 77.08% to 82.92% of the 240 responses they considered. Again, these reliability estimates were considered moderate, and seemed to be influenced by unclearness of responses by the middle age range.

The moderate reliabilities obtained for this form are slightly lower than those reported by previous users of the original Kodroff-Roberge

system. For example, using Cohen's method, Kodroff and Roberge (1975) reported a reliability of .86 for a pair of raters. However, their data were from children who were younger than those in the present studies (first to third grades). The moderate reliabilities obtained for the three observers in the present study suggest that researchers using the K-R system with a broad age range should not try to interpret small differences because such differences may not be reliable even if they do approach statistical significance.

D. RELIABILITY OF THE CLASS - LOGIC SCORING SYSTEM

The same raters and tapes were used to study the reliability of the C-L scoring system as in the other studies. Since the numbers involved were small (10 parts for Form PW, and 8 tapes for Form CA), the data from the two were combined. Even so, there were not enough ratings in some of the categories to produce dependable reliability estimates. However, it should be noted that these categories were not used with some consistency by the three raters and although this contributed to the lower correlations in some cases it does show a consistency among the raters. Tables 10 and 11 show the correlations between raters for the frequency with which they rated categories (with tape as the unit of analysis (n=18)). Also shown in the tables are the number of times both raters in a pair used the category for the same tape.

Generally, it can be seen that the categories that were used more frequently show higher inter-rater correlations, the exceptions being (P=Q) in Table 10 and (P) in Table 11. One can conclude from these data that certain of the categories, notably ($\bar{P}$), ($\bar{P}.\bar{Q}$) and ($P.\bar{Q}$) from the logic ratings, and ($P<Q$), ($P \supset Q$), ($P \subset Q$), and ($P \not\subset Q$) in the class ratings can be used with some confidence of reliability. The other categories should be used with more caution until better reliability estimates can be demonstrated.

Table 10

Correlations between raters for the frequencies that they used class categories, and the number of tapes (n) in which each category was used by both raters for the 18 reliability tapes.[a]

Pairs of Raters	*P=Q*	*P>Q*	*P<Q*	*C* (P Q)	*A*	*T* (P Q)	*E* (P Q)	*G* (P Q)	*O*	*R* (P Q)	*Y* (Q P)	*other*
1 with 2												
r_{xy}	n.s.	n.s.	.87	.69		.50		.77		.76	n.s.	n.s.
n	13	1	11	9		13		8		13	3	2
1 with 3												
r_{xy}	n.s.	n.s.	.87	.87		.68		.97		.89	n.s.	n.s.
n	11	1	11	11		14		8		12	2	4
2 with 3												
r_{xy}	n.s.	n.s.	.77	.90		n.s.		.74		.83	n.s.	.80
n	9	0	9	9		13		7		11	6	14

[a] Note: $p < .05$ for significant correlations.

Table 11

Correlations between raters for the frequencies that they used logic categories, and the number of tapes (n) in which each category was used by both raters for the 18 reliability tapes.[a]

Pairs of Raters	*P*	*Q*	$\bar{P}$	*C* $\bar{Q}$	*A*	*T* $P.Q$	*E* $\bar{P}.Q$	*G* $\bar{P}.\bar{Q}$	*O* $P.\bar{Q}$	*R*	*Y*	$N(P.\bar{Q})$	$P \rightarrow R$ $R \rightarrow Q$	*other*
1 with 2														
r_{xy}	n.s.	n.s.	.92	n.s.		.53	.59	.86	.67			.52	n.s.	n.s.
n	4	3	2	0		18	9	11	12			2	0	2
1 with 3														
r_{xy}	n.s.	n.s.	.83	n.s.		.51	.97	.92	.93			.56	n.s.	n.s.
n	10	5	2	1		18	12	12	10			4	0	3
2 with 3														
r_{xy}	n.s.	.56	.79	n.s.		.86	.66	.83	.68			n.s.	n.s.	.71
n	4	4	1	0		18	12	8	11			2	0	9

CONCLUSIONS AND DISCUSSION

The procedures outlined in this paper represent a fairly complex set of perspectives for the researcher to use in assessing children's reasoning. The procedures are not meant to be comprehensive. However, they do comprise a system which assesses verbal expression (K-R), understanding of implication (C-L), and ability to reason logically (total score). It also assesses understanding of parts of conditional arguments, the arguments presented as a whole, and understanding of the class inclusion aspects of implication. Further, it has been used successfully with children ranging in age from five to 15 years old.

The reliability information provided here was computed on the basis of certain assumptions about how the instruments would be used, and what they would be used to do. For example, the reliability study for the C-L scoring system dealt with the frequency with which categories were rated in the whole protocol because it was assumed that this might be a common research use since the more frequently a category is rated, the more likely that category is part of the child's understanding. However, it is conceivable that researchers might be interested in the frequency with which a category is used when a certain type of argument is presented. Such a change would require that reliability be demonstrated for the new purpose.

The validity of the procedures is shown to some extent by the moderate to high reliabilities demonstrated in this report. Further, the procedures were all tested with children in the age ranged studied, prior to being put together into the total system. In addition, they were based on research and other systems mentioned earlier. For example, the response format (YES, NO, and MAYBE) has been used extensively by other researchers (see Ennis & Paulus, 1965, for example), and the format of the items is widely accepted in tests of reasoning (see Roberge, 1971). However, because the data that are yielded by the system are so diverse and may be used in different ways, it is recommended that a researcher using the system test the validity of his or her data for the purpose they are used.

Finally, the range of reliabilities obtained here suggest that for some of the procedures, small differences should not be relied upon, even if they are statistically significant. This is particularly true for the rating of

interviews. The comment by the observers, which was not systematically investigated for the present report, that responses from children in the third to sixth grade were more difficult to score, should be kept in mind. These procedures might be more reliable if they are used only with younger children or adolescents.

REFERENCES

Antonak, R.F., & Roberge, J.J. Characteristics of the conditional reasoning of educable retardates. American Journal of Mental Deficiency, 1978, 15, 519-528.

Braine, M.D. On the relation between the natural logic of reasoning and standard logic. Psychological Review, 1978, 85, 1-21.

Brainerd, C.J. Piaget's theory of intelligence. Englewood Cliffs, N.J.: Prentice-Hall, 1978.

Bynum, T.W., Thomas, J.A. & Weitz, L.J. Truth-functional logic in formal operational thinking. Inhelder and Piaget's evidence. Developmental Psychology, 1972, 7, 129-132.

Cohen, J. A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 1960, 20, 37-46.

Cronbach, L.J. Essentials of psychological testing. New York: Harper & Row, 1960.

Ennis, R.H. Children's ability to handle Piaget's propositional logic: a conceptual critique. Review of Educational Research, 1975, 45, 1-41.

Ennis, R.H., & Paulus, D. Critical thinking in grades 1-12, Phase 1: Deductive reasoning in adolescence. Ithaca, N.Y.: Cornell University, 1965 (ERIC Document Reproduction Service No. ED 003 818).

Falmagne, R.J. Deductive processes in children. In R.J. Falmagne (Ed.) Reasoning: Representation and process. New York: Wiley, 1975.

Flavell, J.H. Cognitive development. Englewood Cliffs, N.J.: Prentice-Hall, 1977.

Hadar, N.B. Children's conditional reasoning: An investigation of fifth graders ability to learn to distinguish between valid and fallacious inferences. Unpublished doctoral dissertation, Stanford University, 1975 (ERIC Document Reproduction Service No. Ed 118 359).

Henle, M. On the relation between logic and thinking. Psychological Review, 1962, 69, 366-378.

Inhelder, B., & Piaget, J. The early growth of logic in the child. London: Routledge & Keegan Paul, 1964.

Inhelder, B., & Piaget, J. The growth of logical thinking from childhood to adolescence. New York: Basic Books, 1958.

Kodroff, J.K., & Roberge, J.J. Developmental analysis of the conditional reasoning abilities of primary-grade children. Developmental Psychology, 1975, 11, 21-28.

Kuhn, D. Conditional reasoning in children. Developmental Psychology, 1977, 13. 342-353.

Light, R.J. Measures of response agreement for qualitative data: some generalizations and alternatives. Psychological Bulletin, 1971, 76, 365-377.

Nolen, P.A. Implications of formal operational thinking at the college level. Contemporary Educational Psychology, 1976, 1, 269-273.

Nuffield Foundation, Logic. London: Newgate Press, 1972.

Osherson, D.N. Logical abilities in children. Vol. 3. Reasoning in adolescents: Deductive Inference. Hillsdale, N.J.: Erlbaum, 1975.

Parsons, C. Inhelder and Piaget's The growth of logical thinking, II.: A logicians viewpoint. British Journal of Psychology, 1960, 51, 75-84.

Piaget, J. Intellectual evolution from adolescence to adulthood. Human Development, 1972, 15, 1-12.

Piaget, J. The psychology of intelligence. Totowa, N.J.: Littlefield Adams, 1969.

Piaget, J. Judgment and reasoning in the child. London: Routledge & Keegan Paul, 1928.

Roberge, J.J. Analysis of response patterns for conditional reasoning schemes. Psychonomic Science, 1971, 22, 338-339.

Roberge, J.J. A study of children's abilities to reason with basic principles of deductive reasoning. American Educational Research Journal, 1970, 7, 583-596.

Roberge, J.J., & Mason, E.J. Effects of negation on adolescents class and conditional reasoning abilities. Journal of General Psychology, 1978. 98, 187-195.

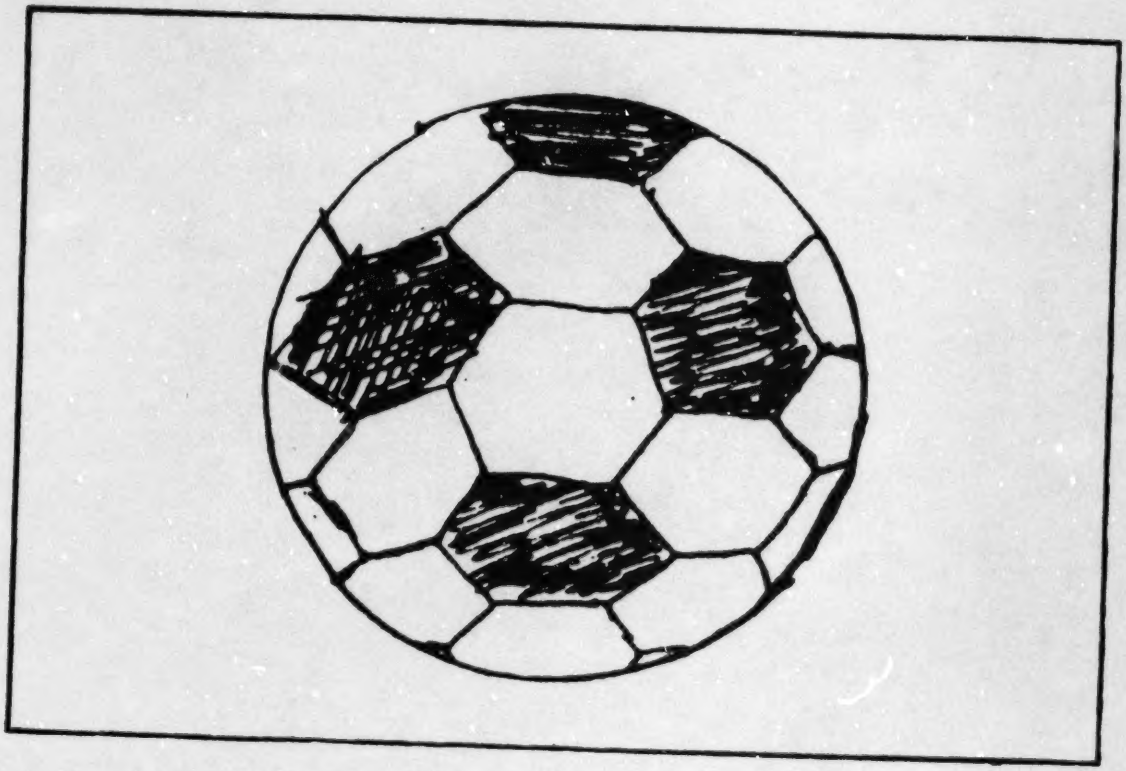
Robinson, W.S. The statistical measurement of agreement. American Sociological Review, 1957, 22, 17-27.

Shapiro, B.J., & O'Brien, T.C. Logical thinking in children ages six through thirteen. Child Development, 1970, 41, 823-829.

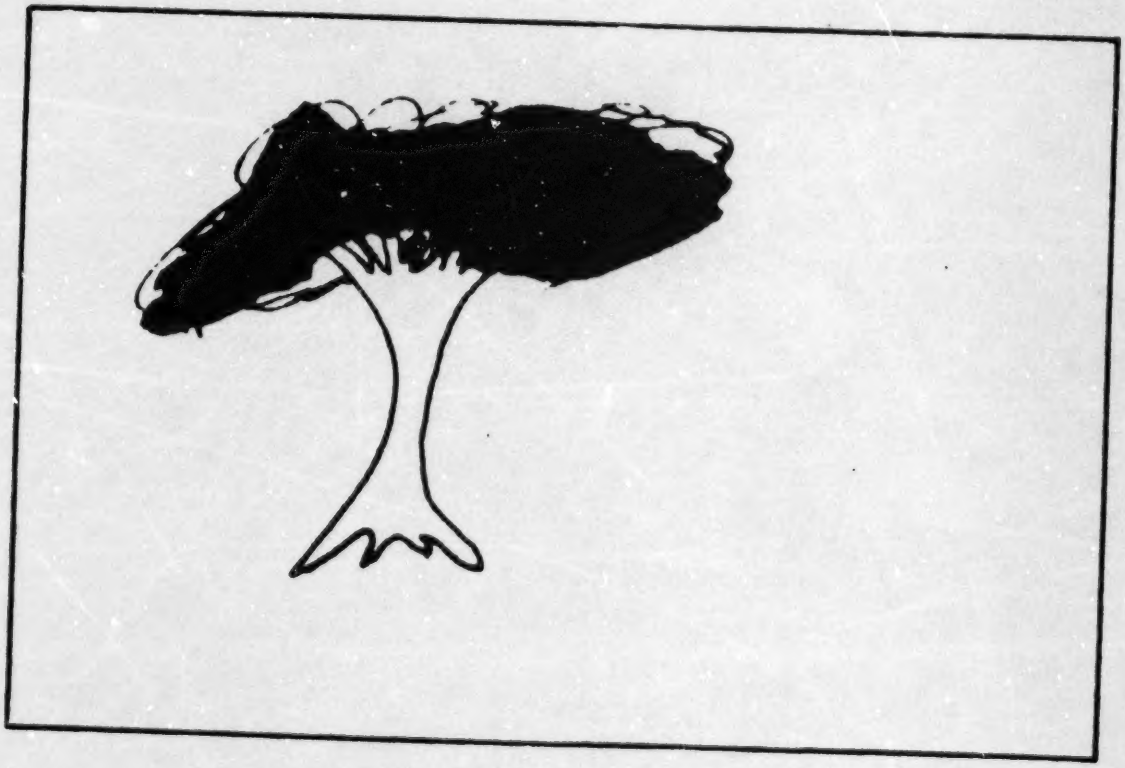
Weitz, L.J., Bynum, T.W., Thomas, J.A., & Steger, J.A. Piaget's system of binary operationals. An empirical investigation. Journal of Genetic Psychology, 1973, 123, 279-284.

BEST COPY AVAILABLE

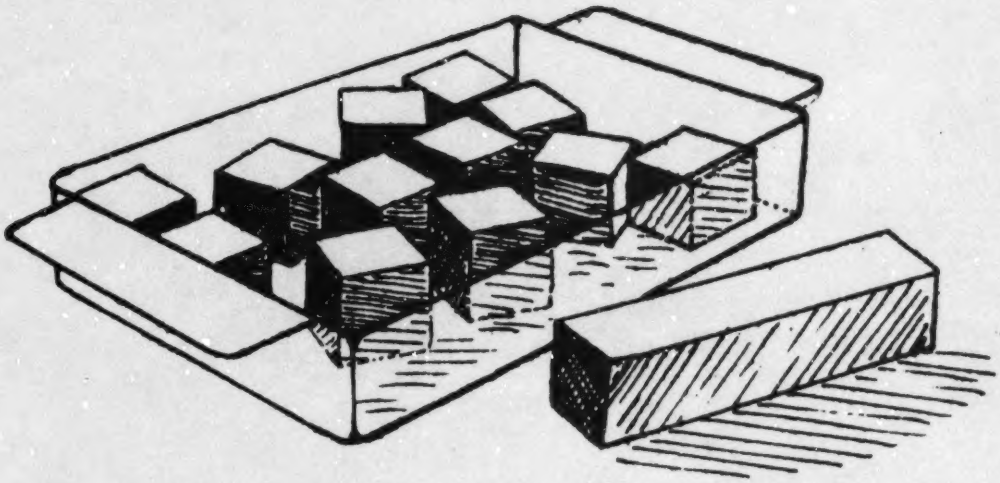
1b

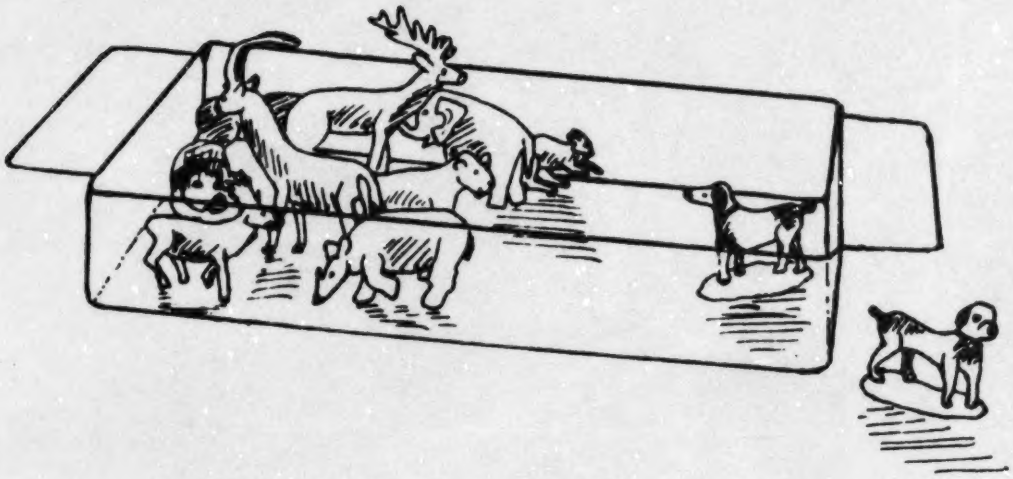


1c



BEST COPY AVAILABLE





BEST COPY AVAILABLE

INSTRUCTIES AAN DE PROEFPERSONEN

Ik ga je een paar dingen zeggen, zoals bijvoorbeeld:

DIT IS MARIA

en dan zal ik iets over haar vertellen zoals: ze heeft een blauwe jurk aan.

Ook zal ik je er direct een paar vragen over stellen. Jij moet er achter zien te komen of het klopt wat ik zeg.

Je kunt op drie manieren antwoord geven:

1) JA. DAT BETEKENT DAT HET MOET KLOPPEN.

2) NEE. DAT BETEKENT DAT HET NIET KAN KLOPPEN.

3) MISSCHIEN. DAT BETEKENT DAT HET SOMS MISSCHIEN WEL KLOPT, MAAR DAT IK JE NIET GENOEG VERTELD HEB OM HET ZEKER TE WETEN.

Bijvoorbeeld, stel je voor dat ik zeg:

MARIA IS (BEVINDT ZICH) NAAST DE KNOOP.

DAN IS (BEVINDT ZICH) DE KNOOP NAAST MARIA.

Is het juist als ik dan zeg dat de knoop naast Maria is?

(Wacht op het antwoord van het kind, zeg daarna:)

Ja is het goede antwoord.

Laten we een ander voorbeeld proberen:

EEN KOE IS GROTER DAN EEN KAT.

DAN IS EEN KAT GROTER DAN EEN KOE.

Is het juist als ik dat zeg?

(Wacht op het antwoord van het kind, zeg daarna:)

Nee is het goede antwoord omdat een koe niet groter en kleiner kan zijn dan een kat.

Laten we nog een ander voorbeeld proberen:

KAAS IS VOEDSEL.

DAN IS VOEDSEL KAAS.

BEST COPY AVAILABLE

Klopt het als ik zeg dat voedsel kaas is?

(Wacht op het antwoord van het kind, zeg daarna:)

Denk er aan dat je alleen JA, NEE of MISSCHIEN kunt zeggen.

(Wacht op het antwoord van het kind, zeg daarna:)

Het antwoord hier is MISSCHIEN omdat voedsel kaas kan zijn of een ander ding dat je kunt eten zoals brood of vlees. Begrijp je dat?

Als het kind ja zegt, vraag hem dan om uitleg.

Zegt het kind nee, leg het dan daarna nog eens uit en gebruik daarbij plaatjes.

BEST COPY AVAILABLE

1. De proefleider zet bak A voor het kind neer en zegt:

ALS ALLE HONDEN DIEREN ZIJN (terwijl de woorden uitgesproken worden, wijst de proefleider op de hond en dan op de rest van de dieren).

EN DIT IS EEN HOND. (de proefleider laat het beeldje van de hond zien die buiten het gezichtsveld was gehouden).

DAN MOET IK DEZE BIJ DE DIEREN ZETTEN (de proefleider zet de hond in de bak).

A. KLOPT DAT?

Als het kind geen standaard antwoord geeft, (ja, nee, misschien) zegt proefleider:

DENK ERAAN DAT JE ALLEEN JA KUNT ZEGGEN. DAT BETEKENT DAT HET MOET KLOPPEN. NEE, DAT BETEKENT DAT HET NIET KAN KLOPPEN. OF-MISSCHIEN, DAT BETEKENT DAT HET SOMS KAN KLOPPEN, MAAR DAT IK JE NIET GENOEG VERTELD HEB OM ER ZEKER VAN TE ZIJN.

Wanneer het kind een standaard antwoord geeft, vraagt de proefleider:

B. WAAROM ZEG JE DAT?

Na het antwoord van het kind, gaat de proefleider door:

C. VERTEL ME WAT HET BETEKENT ALS IK TEGEN JE ZEG, ALLE HONDEN ZIJN DIEREN.

Wanneer het kind antwoord geeft, vraagt de proefleider:

D. ZIJN ER MEER HONDEN OF DIEREN? Dan vraagt de proefleider: WAAROM?

Wanneer het kind antwoord geeft, gaat de proefleider door: VERTEL ME NU WAT DIT BETEKENT:

E. DIT IS EEN HOND (De proefleider wijst op de hond in de bak) DAN MOET IK DEZE BIJ DE DIEREN ZETTEN.

BEST COPY AVAILABLE

1. De proefleider plaatst een blad voor het kind en legt erop:

9 groene kubussen (bij elkaar)

Hij zegt dan: ALS ALLE BLOKKEN VAN HOUT ZIJN (Pl. laat op blad geplaatste blokken zien)
EN DIT IS NIET VAN HOUT (Pl. laat groene kubus zien)
DAN IS HET EEN BLOK (Pl. plaatst groene kubus buiten blad)

A.....KLOPT DAT?

Opmerking 1: Als het kind geen standaard-antwoord geeft, zegt de proefleider: DENK ERAAN, JE KUNT ALLEEN MAAR JA ZEGGEN, DAT BETEKENT DAT HET MOET KLOPPEN, OF NEE, DAT BETEKENT DAT HET NIET KAN KLOPPEN, OF MISSCHIEN, DAT BETEKENT DAT IK JE NIET GENOEG VERTELD HEB OM ER ZEKER VAN TE ZIJN.

Wanneer het kind een standaard-antwoord heeft gegeven, vraagt de proefleider:

B.....WAAROM ZEG JE DAT?

Opmerking 2: Het kind moet in staat gesteld kunnen worden de proefleider duidelijk aan te tonen dat hij/zij de volledige betekenis van de implicatie in de premissen en de conclusie beschreven heeft zoals hij/zij die begrijpt. De proefleider moet het kind blijven ondervragen m.b.t. het antwoord totdat dat duidelijk is.

---

2. De proefleider plaatst een blad voor het kind en legt erop:

8 groene kubussen en 1 gele cirkel (bij elkaar)

Hij zegt dan: ALS ALLE BLOKKEN VAN HOUT ZIJN (Pl. laat op blad geplaatste blokken zien)
EN DIT IS EEN BLOK (Pl. laat gele cirkel zien)
DAN IS HET NIET VAN HOUT (Pl. plaatst gele cirkel buiten blad)

A.....KLOPT DAT?

Als het kind geen standaard-antwoord geeft, zie Opmerking 1.

Wanneer het kind een standaard-antwoord heeft gegeven, vraagt de proefleider:

B.....WAAROM ZEG JE DAT?

BEST COPY AVAILABLE

DEEL 1

Presentatie: Geheel-Concreet

De proefleider heeft twee bakken. De éne bak (A) bevat 10 verschillende kleine beeldjes van dieren (waaronder één hond) die at random neergezet zijn. De andere bak (B) bevat verscheidene groene kubussen van 2.5 cm.
Er is één hond en een niet-gekleurd blokje van 2.5x2.5x8 cm die buiten het gezichtsveld van het kind worden gehouden.

De éne hond in de bak komt apart te staan van de andere 9 dieren.

De mogelijkheid bestaat dat het kind, vooral in de jongere leeftijdsgroepen, niet weet wat het woord 'kubus' betekent. De proefleider moet er daarom naar vragen en eventueel uitleggen.

Wanneer je het kind een object (hond, kubus) laat zien en je ontkent het (Dit is geen hond, geen kubus), bedoel je: laten we net doen alsof.

De proefleider moet het kind ertoe aanzetten redenen te geven. Acceptabele vragen die gebruikt kunnen worden om het kind aan te sporen zijn bijv.:
"daarom zeg je dat?"
"Kun je nu er meer over vertellen?"
"Wil je dat verder uitleggen?"
"Wat bedoel je daarmee?" etc.

Het principe dat de proefleider steeds in de gaten moet houden bij ondervraging van het kind is dat het kind geen aanwijzing gegeven mag worden hoe de vraag beantwoord moet worden.

Indien het kind de instructies in A,B,C,D of E niet begrijpt, herinner het er dan aan wat er gevraagd wordt.
Bijvoorbeeld: Indien de proefleider een reden verzucht en het kind schijnt te antwoorden m.b.t. de validiteit (of het klopt of niet), dan moet het ertoe aangespoord worden om het gevraagde type antwoord te geven.
Let er ook hier op dat het kind geen aanwijzing gegeven wordt m.b.t. het correcte antwoord.

Vermijd, nadat het kind antwoord gegeven heeft, uitspraken als: "Dat is een goed antwoord". Beter is niet zozeer een uitspraak te doen m.b.t. de juistheid van het antwoord, maar meer gericht op het feit dat het kind goed zijn best doet, zoals bijv. "Je doet het goed", "Mooi zo!", ed.

BEST COPY AVAILABLE

1.Enkele vragen om meer te weten te komen over het antwoord van het kind:

Wil je dat verder uitleggen?

Vertel er eens meer over.

Kun je me er meer over vertellen?

Waarom zeg je dat?

Wat bedoel je?

2.Te antwoorden op"ik weet het niet"

a)Herhaal de vraag

b)Weet je wat P is? Qis?

indien "nee"-stop!

indien "ja" - herhaal de vraag

c)"Ik weet het niet"- ga verder met het volgende item.

BEST COPY AVAILABLE

INSTRUCTIES AAN D. PROEFLEIDER

1) Vermijd, nadat het kind antwoord gegeven heeft, uitspraken als: "Dat is een goed antwoord". Geen uitspraak dus m.b.t. de (on)juistheid van het antwoord, maar een uitspraak gericht op het feit dat het kind goed zijn best doet, zoals bijv. "Je doet het goed", "Mooi zo!", ed.

---

2) Het principe dat de proefleider steeds in de gaten moet houden bij ondervraging van het kind is dat het geen aanwijzing gegeven mag worden hoe de vraag beantwoord moet worden. Indien het kind de instructies in A of B niet begrijpt, herinner het er dan aan wat er gevraagd wordt.

---

3) De proefleider moet het kind ertoe aansetten redenen te geven. Enkele vragen om meer te weten te komen over het antwoord van het kind:
"Wil je dat verder uitleggen?", "Vertel er eens meer over.", "Waarom zeg je dat precies?" of "Wat bedoel je daarmee?".

---

4) Wat er gedaan moet worden na het antwoord "Ik weet het niet.":

a) herhaal de vraag; indien hetzelfde antwoord:

b) "Weet je wat P is?", "Weet je wat Q is?"; indien "Nee.": stop met de vraag te stellen; indien "Ja.": herhaal de vraag nog een keer; indien "Ik weet het niet.";

c) stop met de vraag te stellen en ga verder met het volgende item.

BEST COPY AVAILABLE

tape_____ Name_____ Grade_____ School_____ Examiner_____ Scorer_____ FORM CA

egories of lanation	ITEMS 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
bal ficits																																
ory ficits																																
itrary planations																																
uctural ficits																																
ditional asoning																																
terned planations																																
ers																																
d Logic																																

K-R Coding Form.

Place response to syllogism (yes, no, maybe) in box corresponding to explanation, and circle correct responses.

SCORING INSTRUCTIONS FOR
CODING FORM C-L

1. In general, the scorer will place a check in the boxes where the kind of statement the child made is represented at the top of the column. For example, if the child says, "Dogs and animals are the same thing.", checks should be placed in the P=Q column and the PQ column. Also, a check would be place in the P•Q column because the child recogizes that both P and Q can exist at the same time. The specific criteria to use with scoring each column are listed below. The only violation to the check placement in a column is that if the child reverses P•Q in his verbal response, place an R in the appropriate columns instead of a check. For example, if the major premise says, "All P are Q" and the child says "There are Qs and there are Ps". Since the child said that Q and P exist in his answer, we score it as P•Q, and place an R in the column because of the reversal. More than one column can be scored (✓or R) for each item.

2. The three sections of each item that have several-word responses are scored. Each section has a line on the form. There are two forms, one for Parts I and III, and the other for II and IV items.

3. The meanings of the symbols in each column is as follows:

A. Class Relationships

a) P=Q : Response indicates P and Q are the same in number or size.
ex: "They are the same". "There are dogs and there are animals".

b) P<Q : There are more Q than P.

c) P>Q : There are more P than Q.

d) (P)(Q) : P and Q exist as separate variables.
ex: "There are dogs and there are animals". "A thing is either a tree or a plant".

e) (PQ) : P and Q are the same.
ex: "All dogs are animals and all animals are dogs".

f) (P(Q)) : There is some overlap between P and Q.
ex: "Some dogs are animals". "Some animals are dogs".

g) (P(Q)) : P exists as a subclass of Q.
ex: "A dog is an animal". "All dogs are animals".

BEST COPY AVAILABLE

h) (Q P) : Q is a subordinate group to P.

ex: "All animals are dogs","An animal is a dog".

i) Other: Put a symbolfor an other kind of response that needs to be recorded. In the notes area below, the symbol should be explained. Ex:"There are dogs". This would be scored P only.

B. Logical Relationships

a) P : only the existence of P is recognized.

ex:"There is a dog".

b) Q : only the existence of Q is recognized.

ex: "There are plants".

c) $\overline{P}$: only the existence of the negative of P is recognized.

ex: "There is no dog". "This not a ball".

d) $\overline{Q}$: only the existence of the negation of Q is recognized.

ex : "There is no animal."

e) P.Q : The existence of both P and Q is recognized.

ex : "There are dogs and there are animals.", "You said all dogs are animals", "Animals can be dogs and dogs can be animals."

f) $\overline{P}$.Q : recognizes that the negation of P can exist while Q exist.

ex : "There is no dog, but there are animals", "All dogs are animals, but not all animals are dogs."

It can be inferred from some answers that P Q is intended.

ex : "Every dog must be an animal too.", implies that animal is superordinate, and animals can exist that are not dogs.

g) $\overline{P}.\overline{Q}$: recognizes that the negation of P and Q can exist at the same time.

ex : "There are no dogs or animals."

This answer can also be inferred from correctly applied implication or other answers in which P.Q is not stated directly.

ex :"All dogs are animals, but not all animals are dogs.", "If this is a dog, than it must be an animal, because all dogs are animals.", "If this were not a dog, it might or might not be an animal."

h) $(P.\overline{Q})$: says that P can exist even when Q does not.

ex : "There are no round things, but this is a ball anyway.", "This is a dog but not an animal."

BEST COPY AVAILABLE

i) $N(P.\bar{Q})$: denies that P and $\bar{Q}$ can exist at the same time.

ex: "There are no dogs that are not animals".

This can sometimes be inferred from a good implication response, ex:

"This dog must be an animal because all dogs are animals."

j) $\begin{matrix} P \rightarrow R \\ R \rightarrow S \end{matrix}$ means that a term was added. For example,"all trees are green and all plants are green so this must be a plant."

k) Other: Mark any other kind of response that should be recorded (dealing with logical understanding) with a symbol and explain the symbol at the bottom of the page under "notes".

BEST COPY AVAILABLE

-FORM CA-

C-L Coding Form

Name_____ Examiner_____

Tape_____ Scorer_____

Date scored_____

ITEMS	A. Class relationships											B. Logical Relationships											
	P=Q	P>Q	P<Q	P Q	P Q	P Q	P Q	P Q	P Q	Q P	other	P	Q	$\bar{P}$	$\bar{Q}$	P•Q	$\bar{P}$•Q	P•$\bar{Q}$	$\bar{P}$•$\bar{Q}$	(P•$\bar{Q}$)	N(P•$\bar{Q}$)	P→R R→Q	OTHER
1																							
2																							
3																							
4																							
5																							
6																							
7																							
8																							
9																							
10																							
11																							
12																							
13																							
14																							
15																							
16																							
17																							
18																							
19																							
20																							
21																							
22																							
23																							
24																							
25																							
26																							
27																							
28																							
29																							
30																							
31																							
32																							

-FORM PW-

C-L Coding Form

Name_____ Examiner_____

Tape_____ Scorer_____

Date scored_____ Part I III (circle)

ITEMS	A. Class relationships										B. Logical Relationships											
	P-Q	P>Q	P<Q	P Q	P Q	P Q	P Q	P Q	Q P	other	P	Q	$\bar{P}$	$\bar{Q}$	P•Q	$\bar{P}$•Q	$\bar{P}$•$\bar{Q}$	(P•$\bar{Q}$)	N(P•$\bar{Q}$)	P→R R→Q	OTHER
1B																					
1C																					
1E																					
2 B																					
2 C																					
2 E																					
3 B																					
3 C																					
3 E																					
4 B																					
4 C																					
4 E																					
5 B																					
5 C																					
5 E																					
6 B																					
6 C																					
6 E																					
7 B																					
7 C																					
7 E																					
8 B																					
8 C																					
8 E																					

NOTES: